Chair for Statistics and Empirical Economics
Prof. Regina T. Riphahn, Ph.D.

Exam in Panel and Evaluation Methods
Summer Term 2019

Problem 1 (17.5 points)

You want to estimate the effect of compulsory military service on wages. You have a German cross-sectional data set for 2,000 men containing the following variables:

$wage_i$	(logged) wage of individual i.
$service_i$	=1 if individual i carried out military service; =0 otherwise.
$educ_i$	years of completed schooling of individual i.
$reform_i$	=1 if person i was affected by a reform that introduced compulsory military service; =0 otherwise.

You estimate the following model using ordinary least squares (OLS):

$$wage_i = \beta_1 + \beta_2 service_i + \beta_3 educ_i + \epsilon_i$$

1.1 Your estimates yield, among others, the following results: $\hat{\beta}_2 = 0.3$, with an estimated standard error $se(\hat{\beta}_2) = 0.1$. Interpret the coefficient in terms of economic and statistical significance at the 1% level. (2.5 points)

1.2 Name and explain a reason why the variable $service_i$ could be endogenous in this model. (2 points)

1.3 Give an example for and explain a situation in which the estimated coefficient for *service* could be biased upwards. (2 points)

1.4 Assume that the variables $reform_i$ and $educ_i$ are exogenous. The (Durbin-Wu-) Hausman test can be used to check whether the variable $service_i$ is actually endogenous. Sketch the test procedure including the necessary estimation equations. (4 points)

1.5 Provide the necessary estimation equations for a *two-stage-least-squares* estimation (2SLS) where you instrument the service variable using the variable $reform_i$. (3 points)

1.6 State the moment condition(s) required for IV estimation. In which situation can you use the Sargan test to check the exogeneity of the instrumental variable? (2 points)

1.7 Name two properties that must be met by valid instruments. (2 points)

Problem 2 (5 points)

2.1 Define the Intention to Treat Effect (ITT) verbally and give an example for a situation in which you would attain an ITT as your estimation result. (3 points)

2.2 Show formally how the Average Treatment Effect on the Treated (ATT) can be calculated from the ITT. Define all new variables you introduce. (Definitions of ATT and ITT are not required.) (2 points)

Problem 3 (14 points)

At A-Town university, students receive 5 bonus points for the final econometrics exam if they pass a compulsory trial exam taken in the middle of the semester. You have the following data about the participating students:

$trial\_score_i$	score in the trial exam (between 0 and 30).
$final\_score_i$	score in the final exam (between 0 and 60).

To pass the trial exam, students have to reach a trial score of at least 15 points. To pass the final exam, students have to reach a final score of at least 30 points, including bonus points. You intend to use a regression discontinuity design (RDD) to examine how receiving 5 bonus points from passing the trial exam affects the final exam score.

3.1 Which RD design would be appropriate in this case? Elaborate briefly. (1.5 points)

3.2 State the main identifying assumption of RDD and briefly state two ways to test whether this assumption holds. (3 points)

3.3 Write down the RDD estimation equation to estimate the causal effect of receiving the 5 bonus points on the final exam score. Define new variables if necessary. (2 points)

3.4 When the professor graded the trial exam, she shifted those students who were just below the passing score above the threshold. What does this mean for the validity of your RDD strategy and why? (1.5 points)

3.5 The professor teaches the same course at the university of neighboring B-Town. There she also introduced the trial exam, but the students do not get bonus points for passing. You also have the data from B-town.

3.5.1 Write down a regression model to estimate the causal effect of the 5 bonus points on students' scores in the final exam with a Difference-in-Difference (DiD) estimation. Define new variables if necessary. (You do not need to repeat definitions already given in your answer to problem 3.3.) (3 points)

3.5.2 State and verbally explain the central assumption that has to hold for the DiD method to identify the causal effect of the 5 bonus points. Give an example for a situation in which the assumption would be violated. (3 points)

Problem 4 (11 points)

The government tests a drug abuse intervention program that provides former drug addicts with professional counselling for 6 months after they finish their rehab. Participation in the program is voluntary. The program is only offered in one city. You want to evaluate if the program was successful in reducing participants' relapse probability. For the evaluation you use Radius-Caliper-Matching with a control sample from another city.

4.1 Explain what a caliper is. (2 points)

4.2 What is the trade-off when deciding about the size of the caliper? Elaborate. (3 points)

4.3 Define the Stable Unit Treatment Value Assumption (SUTVA). Construct a short example for a situation in which the SUTVA would be violated in the example of the evaluation of the intervention program. (3 points)

4.4 Define the condition of common support (overlap). Construct a short example for a situation in which this condition would be violated in the example of the evaluation of the intervention program. (3 points)

Problem 5 (14 points)

Using individual-level panel data, you examine how education impacts life satisfaction. Your dataset contains the following variables that were collected over 20 survey waves:

$lsat_i$	Life satisfaction of person i, on a scale from 0 (worst) to 100 (best).
age_i	Age of person i in years.
$educ_i$	Completed years of schooling of person i.
$female_i$	=1 if person i female; =0 otherwise.

The following table shows the results of a linear regression with individual-specific fixed effects:

```
      Source |       SS           df       MS      Number of obs   =   131,610
-------------+----------------------------------   F(3, 131606)    =    626.82
       Model |  539254.575         3  179751.525   Prob > F        =    0.0000
    Residual |  37740212.1   131,606  286.766653   R-squared       =    0.0141
-------------+----------------------------------   Adj R-squared   =    0.0141
       Total |  38279466.7   131,609  290.857515   Root MSE        =    16.934
------------------------------------------------------------------------------
        lsat |      Coef.   Std. Err.      t    P>|t|     [95\% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |    .17817     .04959      3.59   0.000       .0809      .2754
         age |   -.00169     .01157     -0.15   0.884      -.0244      .0209
      female |  -5.59071   11.61245     -0.48   0.630    -28.3509    17.1695
        cons |  78.41587    6.21110     12.63   0.000     66.2422    90.5895
-------------+----------------------------------------------------------------
```

5.1 Interpret the magnitude of the coefficient for education. (1 point)

5.2 Which individuals identify the gender effect in the model? Give an example. Is the estimated coefficient meaningful? (3 points)

5.3 What is a balanced panel? Does this data represent a balanced panel? (2 points)

5.4 You used uncorrected standard errors. Name one disadvantage of this approach compared to cluster-robust standard errors and a consequence for your inference. (2 points)

5.5 Use a Hausman test at the 1% significance level to check if a fixed-effects model or a random-effects model should be used to estimate the model. There are no further explanatory variables available apart from the variables used. Assume an empirical value for the test statistic of 11.44. State the null and alternative hypothesis, the number of degrees of freedom and the critical value. Explain your test decision. (4 points)

5.6 Name two advantages of FE-quantile regression versus FE regression. (2 points)